PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : G06F 15/70, G01C 3/00

A2

(11) International Publication Number: WO 88/ 02518

(43) International Publication Date: 7 April 1988 (07.04.88)

(21) International Application Number: PCT/GB87/00700

(22) International Filing Date: 2 October 1987 (02.10.87)

(31) Priority Application Number: 8623718

(32) Priority Date: 2 October 1986 (02.10.86)

(33) Priority Country: GB

(71) Applicant *(for all designated States except US)*: BRITISH AEROSPACE PUBLIC LIMITED COMPANY [GB/GB]; 11 Strand, London WC2N 5JT (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)* : WRIGHT, Steven, M. [GB/GB]; Department of Engineering, Manufacturing Engineering Group, University of Cambridge, Mill Lane, Cambridge CB2 1RX (GB).

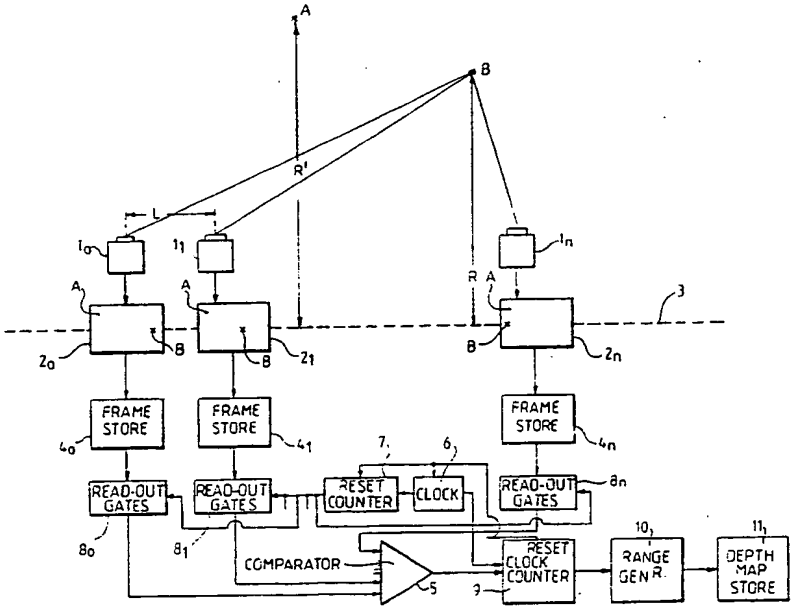
(74) Agent: EASTMOND, John; Corporate Patents Department, British Aerospace PLC, Brooklands Road, Weybridge, Surrey KT13 0SJ (GB).

(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: REAL TIME GENERATION OF STEREO DEPTH MAPS



(57) Abstract

An automatic machining or assembly system including a comparator for comparing the intensity of a pixel in a first image of a scene produced by a sensor with a corresponding pixel and pixels increasingly displaced from the corresponding pixel in a second image of the same scene displaced with respect to said first image and for producing signals representing image depth the magnitude of which is determined by the relative displacement of compared pixels having minimum intensity variation or by a second sensor linearly displaced from the first sensor or by optical diffraction means between the first sensor and the scene, when rotated to a new position.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	ML	Mali
AU	Australia	GA	Gabon	MR	Mauritania
BB	Barbados	GB	United Kingdom	MW	Malawi
BE	Belgium	HU	Hungary	NL	Netherlands
BG	Bulgaria	IT	Italy	NO	Norway
BJ	Benin	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SN	Senegal
CH	Switzerland	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
DE	Germany, Federal Republic of	LU	Luxembourg	TG	Togo
DK	Denmark	MC	Monaco	US	United States of America
FI	Finland	MG	Madagascar		

REAL TIME GENERATION OF STEREO DEPTH MAPS

The present invention relates to industrial vision systems and in particular to the application of such systems to the field of flexible automation of small batch manufacturing and robotic assembly systems.

It is a general requirement of such systems that they should be able to recognize objects being manufactured or assembled, compare this with the desired final configuration of the object and take further machining or assembly actions appropriate to achieve that final objective.

Some known industrial vision systems use a computer to store information relating to the geometry of components to be machined or assembled in a database. This information is compared, during the operation of the automatic machine, with the images obtained by a camera or other suitable sensor. The sensor measures variation in intensity which are then compared with shaded images stored in the computer database and inferences made about the type and position of the objects in the field of view.

The disadvantages of such known systems are that variations in ambient lighting conditions can lead to intensity, and colour variations or surface finish variations, reflections and shadows all of which complicate comparison with the stored database. Inferring object geometry from an variations in image intensity is, therefore, a non- trivial task due to such variations. As an alternative to intensity map comparison of objects and their stored image recent systems under development attempt to compare range variation of object features, hereinafter referred to as 'depth map', with the stored

object parameters. Depth maps may not be affected by ambient lighting variations etc. Ideally, an industrial sensor to be used in conjunction with such a database would, therefore, provide a depth map, independent of spurious light intensity changes, the primary sensor information.

A number of such systems are available or being developed based on active lighting, scanning ultrasonic or laser rangefinders, stereo image analysis or some combination of these techniques. It these is a requirement of the types of such systems which are to be routinely applied in industrial robotic assembly, that the sensor performance must not degrade when analysing "difficult" images such as those containing very repetitive features, because the automatic manufacturing system will be expected to operate unsupervised on a wide variety of tasks.

Ideally the sensor will be self-contained, such that it is possible to mount it on the end of a robot arm so that the viewpoint and the scale of the image are under program control. This has the additional benefit that the dynamic range of such a depth mapping sensor need not be as large as with a fixed camera system since the area requiring maximum depth resolution eg, the object being manufactured will usually be close to the robot end effector, and even when this is not the case, the robot arm may be moved to the area of interest to "take a closer look".

Each of the numerous possible approaches to obtaining a depth map, have advantages in particular applications. For instance, scanning rangefinders can avoid ambient lighting problems but are inherently slow and as precision mechanical devices, are also likely

to remain expensive. Other structured light approaches can process a large number of points in parallel but require setting up for each application.

Stereo image analysis requires that corresponding pixels in the two views are identified. This requires an iterative procedure which complicates implementation in hardware, and can also give ambiguous results in the presence of repetitive features in the image, or if features in the image, are nearly aligned with the axis of the stereo pair.

An alternative technique for determining scene depth is "Range from Motion". This approach can avoid the correspondence problem by tracing features between adjacent images, but if the full six degrees of freedom of motion are allowed, the calculations to recover the scene depth can be complicated, particularly if the relative position of each depth can be complicated, particularly if the relative position of each view is independent motion in the scene during the image sequence capture time.

If the vision system is to be included in any continuous feedback loops then a further requirement will be that its frame rate ie, the rate at which it produces a complete depth map of the object must be fast enough not to limit the dynamic performance of the robot - this will require a depth map calculation delay of less than say 0.1 seconds for a typical modern robot such as the IBM 7565. This will almost certainly require a dedicated analysis microprocessor. If this is to be conveniently achieved, then the analysis algorithm controlling the microprocessor operation must be as simple as possible.

An object of the present invention is to provide industrial vision system apparatus incorporating a technique for constraining the "range from motion" problem in order to simplify subsequent analysis while still meeting the above requirements for the sensor to be self-contained, and able to deal with "difficult" images.

Another object of the present invention is to provide an industrial vision system which does not require any significant manual setting-up for a new task, is insensitive to changes in ambient lighting, processes information at a rate matched to the dynamic performance of the machining or assembly tool and is not unduly expensive.

It is a further object of the invention to provide an industrial vision system which recognizes the position and orientation of an object to be machined or assembled, provides dynamic feed back to control the machine or assembly tool relative to datums in the visual field without the need to provide complex jigs and fixtures and with automatic inspection of components and assemblies in comparison with with a stored specification.

According to the present invention an automatic machining or assembly system includes a comparator for comparing the intensity of a "pixel" in a first image of a scene produced by a sensor with a corresponding pixel and pixels increasingly displaced from the corresponding pixel on a second image of the same scene displaced with respect to said first image and for producing signals representing image depth the magnitude of which is determined by the relative displacement of compared pixels having minimum intensity variation. The first and second images may be produced by first and

second sensors respectively of a plurality of at least three sensors linearly displaced from each other or by rotatable optical diffraction means between the sensor and the scene, when rotated to first and second positions.

According to the invention in another aspect thereof an industrial vision system includes either at least three spaced-apart sensors each sequentially and synchronously producing a separate one of a corresponding at least three images of a scene, or a single sensor and means for periodically varying its field of view to produce said at least three images frame storage means for storing data representing the intensity of a plurality elements of each image as it is produced and comparatory means for sequentially comparing the intensity of an element of an image produced by one sensor with selected elements of an image produced by another sensor and for producing a range signal the magnitude of which is determined by the relative displacement of the elements of each image giving rise to the minimum intensity variation there-between.

Preferably the system includes a regular geometric array of three or more sensors and the comparator sequentially compares intensities of pixels in the images produced by the sensors relative to a datum pixel in the image from a datum sensor adjusted by amounts directly proportional to the position of the sensor in question relative to the datum sensor. The use of at least three sensors in such an array overcomes the potential ambiguities that a two sensor system would give rise to repetitive features in the scene.

The sensors may be conventional television cameras.

A passive array of well matched television cameras (or alternatively a single camera rapidly scanning through a fixed

sequence of positions) can give the advantages of range from motion in avoiding any ambiguity in pixel correspondence. The virtually simultaneous image acquisition (or short sequence capture time) of such cameras minimises the problem of independent movement in the scene. The fixed set of camera positions simplifies calibration by removing the reliance on a separate motion sensing system, and careful choice of camera positions can then minimise the complexity of the subsequent calculations and eliminate the sensitivity to feature orientation. In particular, the degrees of freedom of viewpoint position can be reduced from six to two if the cameras are arranged in a planar array normal to their collective line of sight.

In the short term, an array of television cameras would be expensive, bulky, and difficult to calibrate, and an alternative technique uses the variable parallax offset produced by a rotatable block of perspex to scan a sequence of viewpoints on to a single camera. Two configurations may then be used; a linear, and a circular scan sequence. These are seen as practical systems in their own right, particularly when combined with a high frame rate CCD camera.

The intensities of pixels in the images produced by the sensors may be encoded in binary form and stored sequentially in frame stores for later bit-by-bit comparison with the appropriate encoded intensities of pixels produced by other sensors. Alternatively the intensities of pixels in each image may first be transformed using, a Hough transform technique, into a sequence of signals each having one of three possible values; a first value ascribed when adjacent pixels have equal intensities; a second value ascribed when the intensity of a pixel is less than that of the previous pixel in the scan sequence;

or a third value ascribed when the intensity of a pixel is more than that of the previous pixel in the scan sequence. The sequence of three value signals in one image is then cross correlated with the sequence of three value signals in another image and the pixel displacement giving rise to minimum intensity variation determined from the maximum correlation signal output.

The invention is perhaps best understood by reference to the analogy of scenes seen by passengers sitting at different windows along the length of a train. If the train is moving, objects in the scene presented to an observer at one window will be displaced in that scene, during any given time interval, by an amount dependent on the range of that object from the window and the observer. Thus objects in the foreground such as sleepers in the adjacent track move rapidly from one side of the scene to the other whilst objects in the distance such as distant mountain ranges hardly move at all or effectively remain at the same point in the scene.

It would indeed be possible as has been mentioned above to construct such a "range from motion" imaging system using a single sensor to track the motion of image points between a sequence of very similar images. However, this approach is limited by the accuracy with which the motion between the images can be measured, by the speed of any independent motion in the scene (vehicles moving relative to the train) and by the complexity of the calculation required to reconstruct a depth map given the possible six degrees of freedom of the original motion. This calculation can also be ill-conditioned, for example, if objects are directly in the line of motion (i.e., the train driver's view). The principle disadvantage of such a system

would be the unacceptable time it would take to construct a complete depth map.

Returning to the analogy, if each of a number of observers on the train sitting at different windows simultaneously measures the position of a particular object in the plane of his window, the co-ordinates of each object relative to each window frame will be displaced by an amount dependent on the window separation and the range of that object. If the object is at the same co-ordinate position in each of the windows then clearly that object is infinitely distant from the observer whilst if the co-ordinates of the position of the object vary a great deal from the window to window the object must be at a relatively short range.

The latter phenomena is utilized in the present invention where the problems in the "range from motion" analysis are avoided by using a camera array or other suitable sensor array to obtain a number of views simultaneously. The relative position of objects in each of the scenes produced by cameras can therefore be established accurately by the use of mechanical constraints and the problem of motion within the scene is minimal. The accuracy of this technique is maximised if the line of sight is normal to the effective motion between views, and its sensitivity to directional features in the scene is minimised if the cameras are placed in a plane rather than in a line. If a displacement of the cameras from each other, i.e., of the view-points, is known then the images may be correlated to reveal the amplitude of the parallax motion and hence the range.

One of the main difficulties with this correlation approach to range measurement is its dependence on image features. Were only two

sensors to be used it would be clearly be difficult to distinguish the relative displacement of one object in its image produced by one sensor from its image produced in the other sensor from the actual displacement of two similar objects producing similar intensities in the images produced by both sensors. The more camera positions there are the greater the possibility of removing this ambiguity becomes. Range estimates are in any case only available where a significant intensity gradient exists in the image otherwise the technique produces a sparse depth map. This situation can be improved by creating intensity gradients in otherwise uniform surfaces by active lighting, and as the objective is only to provide variable intensity profiles this aspect of the system need not be accurately be set-up or expensive.

Embodiments of the invention will now be described with reference to the accompanying drawings of which:-

Figure 1 shows a schematic diagram of a depth-map producing system according to the invention,

Figure 2 is an optical-ray diagram illustrating the relationship between the positon of an image point in a scene, the object point and the camera lens in a 3-dimensional depth-map,

Figure 3 shows a schematic diagram of an alternative depth-map producing system according to the invention,

Figure 4 shows part of a depth-map producing system according to the invention using a single camera sensor,

Figure 5 illustrates the effect of camera spacing in a system according to the invention.

Figure 6 illustrates an experimental set up to demonstrate the operation of the invention and

Figure 7 - 12 show typical results obtained from the set-up of Figure 6.

In Figure 1 a linear array of television cameras 1_0, 1_1,....1_n each with identical sensitivity and each producing equal area images 2_0,2_1,...2_n of a scene including 2 objects A and B where A at an infinite range R' from the image plane 3. The cameras are equi-spaced and separated by a distance L from each other and the object B is at a finite range R from the image plane 3. Each camera is associated with a frame store 4_0, 4_1,....4_n which stores in digital form the intensities of each pixel in the images 2_0, 2_1,...2_n respectively.

A comparator 5 is arranged to compare intensities of certain pixels in corresponding lines of the scan images 2_o,2_1,...2_n. A lock 6 and counter 7 control the operations of read-out gates 8_0, 8_1,...8_n so that intensity values of corresponding pixels, then of pixels displaced by one pixel space, then of pixels displaced by two pixel spacings etc. in the images produced by adjacent cameras are sequentially compared by the comparator 5.

The comparator 5 is arranged to produce an output signal when and only when the variation in pixel intensities compared is zero or a minimum. When this occurs a counter 9, reset to 0 at the start of every comparason cycle and controlled by the clock 6 is read by a range signal generator 10 to produce a signal corresponding to the appropriate object range for storage in a depth-map store 11. The contents of the depth map store may be compared in further apparatus (not shown) with a stored depth map model of an object to be machined

or assembled and control signals produced accordingly to control the machining/assembly actions of a machine.

It will be appreciated that if a particular object is at infinity,such as object A in Figure 1,its image will occupy the same pixel or group of pixels in each of the images $2_0,2_1,\ldots 2_n$ (see the train/passenger analogy above). In this case the very first comparison of intensities made by the comparator 5 reveals zero or minimum intensity variation of the pixels in each of the images stored in the frames store $4_0,4_1,\ldots 4_n$. A range signal is therefore generated by the range generator 10 corresponding to infinite range, infinite depth.

In general an object, such as B, will occupy pixels $B_0,B_1,\ldots B_n$ in images $2_0,2_1,\ldots 2_n$ respectively and the displacement of these pixels relative to the pixels in the adjacent image will depend on the inter-camera spacing and the range R of the point B from the image plane 3. The comparator 5 will thus only produce a signal corresponding to minimum or zero intensity variation when the read-out gates $8_0,8_2,\ldots 8_n$ are controlled to compare intensity values of pixels having the appropriate inter-image displacement. In effect the images $2_0,2_1,\ldots 2_n$ are sequentially overlayed to an ever increasing extent until all the pixels containing the image of the point B are super imposed. At this point comparator 5 produces output signal which stops the counter 9 at a count corresponding the a range R as decoded by the range generator 10 and the range information is fed to the depth-map store 11.

The clock rate of the clock 6 is chosen such that the complete cycle of pixel comparisons in a line or set of lines of the images

$2_0, 2_1, \ldots 2_n$ is completed in a time sufficiently short to enable the complete map to be stored in the store 11 in the period following the complete frame-scan by each of the cameras $1_0, 1_1, \ldots 1_n$

The data resulting from one image capture period can be considered as a four dimensional data solid. The conventional stereo analysis approach would identify features in each individual camera image, then track the motion of these features between images in order to establish the magnitude of the parallax offsets, and hence the range. However the data can be analysed in an orthogonal direction to these camera images for the case of a three dimensional data solid (the data structure which would be obtained from a linear array of cameras).

For the three dimensional case, a section through the data solid orthogonal to the image plane results in a new image (hereafter referred to as the epipolar or orthogonal image) which consists entirely of linear structures. From inspection of Fig 2, it may be seen by similar triangles that the offset of an image point from the centre of an image (Xi) is related to the offset of the object from the camera axis (Xp) by the equation:-

$$Xi = F.Xp/Z \qquad (1)$$

Where F is the focal length of the camera, and (Z+F) is the length of the normal from the object point onto the image plane. The slope of the lines in the orthogonal image space is d(Xi)/d(Xp) it may be seen from (1) above that this slope is inversely proportional to the Z co-ordinate of the object point. In a planar array of cameras, the four dimensional data solid results in an orthogonal image hypersurface where the gradient of the hypersurface is inversely

proportional to range as above. Measurement of the slope of these image features will, therefore, provide a cartesian depth map.

These image coordinates xi could be identical in all the sensors if correction factors of Xi*F/Z are added to the x calculation for each image. Under these circumstances the variance between the intensity values of the pixel xi in each of the test images will be zero. As F is a constant and Xi is known with some accuracy for each sensor, the variance between the set of image intensity values obtained can be plotted as a function of test values of Z, and the range estimated by detecting the minimum in the resulting variance profile. (Note that the values Xi are calculated relative to one of the sensor images for which the correction factors are zero for all Z - i.e. for all test values of Z there is at least one of the test points at the final intensity, therefore it is only at one particular range where the variance can equal zero.)

For real images this process can be repeated for each pixel in the master image to provide a full depth map. A computer algorithm can readily be devised to implement the process. In order to arrive at a data flow description of the chosen algorithm the values of Xi, and the test values of Z need to be chosen such that only integer frame store addresses are needed, and such that the data flow algorithm has a regular structure which is easily mapped into hardware.

The integer addressing requirement can be achieved by choosing the values of Xi to be some integer multiple of a constant (C-Space) and modifying the test value sequence appropriately. The resulting algorithm will then be image position invariant, and is therefore well suited to parallel processing by for example an array processor.

Figure 3 shows a possible implementation of the algorithm as a pipeline, data flow process in which the pixel by pixel intensity data streams 13_0, 13_1, $13_2 \ldots 13_n$ from each of n cameras (not shown) in an array are fed into separate First-In-First-Out (FIFO) buffers stores 14_0, 14_1, $14_2 \ldots 14_n$. The delayed outputs corresponding to a first range R_0 of the buffer stores are added in gates 15_0, 15_1, $15_2 \ldots 15_n$ and the resultant digital signal representing the combined intensity of those pixels is applied to one input of a comparator 16_0. The FIFO buffer stores 14 and adder circuits 15 associated with a range R_0 form an R_0 unit 17. The unit 17 may, for example add the intensities of adjacent pixels from each of the camera images at any given time.

The data streams 13 are simultaneously fed into units 17^{1}, 17^{11}, 17^{111}, etc. associated with ranges R_1, R_2, R_3, etc. Each unit is identical, having the same number of FIFO buffers 14 and adders 15 but in each the delay between pixels added is successively increased. The output signal from unit 17^{1} is compared with the signal from unit 17 in the comparator 16_0. The minimum signal of the two is then compared in comparator 16_1, with the signal from unit 17^{11} and so on. The output from the final comparator circuit 16_m (if m ranges are considered) represents the signal from the unit 17 corresponding to the range at which there is minimum variation in the intensity of pixels compared and hence corresponds to a particular range R of that point in the scene viewed by a datum camera.

The hardware required for the analysis can be considerably simplified by allowing only linear arrays of sensors as the parallax offsets being measured can be arranged to be parallel to the camera raster scan row direction. A number of these linear elements can be

combined to form a 2D array of sensors this allows the FIFO buffers to be replaced by a series of "D type" registers. A further simplification is possibly if the image can be thresholded to reduce the number of bits per pixel as this would simplify the difference, addition, and comparison operations. In the extreme case the significant edges can be detected in the images by for instances zero crossing analysis, and these binary images cross correlated to detect a sparse range map.

In the case of a linear array of cameras arranged normal to the line of sight, the extreme camera separation defines the accuracy obtainable as it would for a normal stereo pair. The intermediate cameras establish the correspondence of pixels in the two views in the presence of repetitive features in the image.

In order to unambiguously establish the correspondence of pixels in the additional image of the scene with pixels already observed by a camera or set of cameras, the actual position of a projection onto the new image of a scene point must lie within a tolerance band of the position expected from the image set already available. The tolerance band is defined by the wavelength (L) of the highest spatial frequency present in the image. This is produced by the more repetitive feature in the scene when it is at the minimum range of interest (zmin). It can be shown that the spacing (Bn) of the "n-th" camera from its predecessor is given by:-

$$Bn = (Zmin * L_n) / F \quad \ldots\ldots\ldots\ldots 2$$

where F is the focal length of the camera.

The maximum spatial frequency which can be detected has a wavelength of twice the pixel spacing (S). Using this value the series of camera spacings becomes:-

$$Bn = (Zmin * (2 * S)^{n}) / F \quad \ldots\ldots\ldots 3$$

The equations governing depth quantisation of a stereo pair of cameras are:-

$$Zn = (B * F) / (N * S) \quad \ldots\ldots\ldots\ldots\ldots 4$$

where N = an integer displacement of the image between the two cameras

B = baseline of the stereo pair

S = pixel spacing

F = Focal length

Zn = depth

The separation of the range quantisation levels is given by:-

$$(Zn - Zn-1) = B* F * [(1-N)/(N+1)]/(N*s)$$

$$= Zn (1/(N + 1)) \quad \ldots\ldots\ldots\ldots\ldots 5$$

If say P% range resolution is required over a range of interval Zmin to Zmax then the maximum allowable separation of quantisation levels (MS) is given by:-

$$MS = Zmax * P / 100$$

$$= Zn [1 / (N + 1)]$$

therefore

$$N = (100 / p) - 1$$

but $Zmax = B * F / (N * S)$

therefore

$$B * F = [(100 / P) - 1] * Zmax * S \quad \ldots\ldots\ldots 6$$

This relationship defines the maximum baseline required as a function of the camera focal length, resolution, and the requirements for accuracy and range. A possibility opened up by the use of a camera array is that the camera spacing can be constant increments, and the focal length varied to satisfy equation 6 rather than the constant

focal length approach which is a more direct extrapolation of the 'range from motion' algorithms. This results in a more compact sensor, but the different scale of the pixels in each view could complicate the analysis by, for instance requiring a rolling average filter of long focal length to maintain the scale.

A typical requirement in a robotic assembly might be:-

Zmax = 2.0 meters

Zmin = 0.5 meters

P = 5

Camera Resolution = 512 by 512 pixels.

In a planar array the separation of the outermost cameras defines the accuracy with which the object range can be determined. The range quantisation levels Z (n) for this camera pair can be obtained by substituting into equation (1) for the maximum baseline (B) and the image offset (Xi = n.S where S is the pixel spacing, and n is a positive integer) to give:-

$$Z(n) = (F.B)/(n.S) \quad (n=1,2,.....) \qquad (7)$$

This equation allows calculation of the furthest detectable range by evaluation of (7) at n=1. These calculations provide a guide to the size of array needed to fulfil a given requirement, but are not a theoretical limit to its performance, as it may be possible to locate edges to sub pixel resolution in a process similar to hyperacuity in human vision. It should be noted that the inverse law governing the separation of quantisation levels implies that accurate range profiles are available for a very restricted range of depths. In a practical system, if a scanning perspex block 40 is used as shown in Figure 4 in conjunction with a single camera 41 this range of fine sensitivity can

be adjusted during operation by use of a zoom lens, as well as by changing the camera position.

Intermediate cameras in the array maintain the correspondence of pixels between any two views. This can be illustrated by considering the case of a linear array of cameras and a test object with regular vertical lines at a fixed range. This arrangement would produce the orthogonal image shown in Fig 5. If the cameras spacing is increased to an interval such that the image is displaced by one wavelength or greater between successive samples, then the mapping into the parameter space will produce ambiguous (or aliased) results as illustrated by the set of samples marked # in Fig 5. In general, for an extra camera on the end of a line of 'N' cameras the actual position of an image feature in the additional image must not deviate by more than one wavelength of the highest spacial frequency present in the image from its expected position. This observation provides a theoretically optimum number of cameras for any given accuracy and resolution based on an exponentially increasing camera spacing. In practice this theoretical optimum can only be approximately achieved as the use of one vote to distinguish between alternative maxima in parameter space makes no allowance for sensor noise and inaccuracy.

Instead of digitising the intensity values of each pixel for comparison as shown in the apparatus of Figures 1 and 3 a Hough Transform technique may be used.

The Hough Transform is a mapping from image space into parameter space, which was originally developed to identify the parametric form of straight line features in images, and has since been extended to analytic curves, arbitrary shapes, and can be applied to multi

dimensional data solids described above in order to find the slope of features in the orthogonal image set. Returning to Figure 5 it will be appreciated that the Hough Transform would accumulate more votes for the correct range line (b) than for the aliased lines (a and c).

It may be seen from equation (1) that the features in the orthogonal image space corresponding to short ranges in the parameter space may, if they are already close to the edge of that solid, exceed the boundary of the image solid without appearing in all the images. It may also be that they are interrupted by an occluding object. Noise in the image, and alignment or sensor matching problems may cause the vote lines in parameter space to cluster rather than intersect at exactly one point for every object point.

One of the difficulties presented by a pipeline architecture such as that shown in Figure 3 running at video rate is that of screen wrap-around where the pixel for the extreme right of the picture is followed through the pipeline by the extreme left pixel from the line below. The range estimation algorithm involves simultaneous comparison of pixels at different x offsets in the data streams from a number of cameras, and the calculation must therefore be inhibited at the end of lines. This can be done in a number of ways;-

1) The calculation of all the intensity variance estimates can be terminated once any one of them reaches the end of the line.

2) The calculation of variances corresponding to the shorter ranges require a larger offset than the longer ranges. The

field of view given by option 1) above can therefore be extended but with progressively more limited range quantisation by only inhibiting those range calculations which have reached the end of line.

3) The performance of option 2) can be further extended as each range calculation can proceed even if it normally requires data from beyond the end of line, by assigning 'wild card' values to these data elements. This gives a full set of range estimates over the full field of view, but based on a reducing number of cameras, and so the shorter range estimates would have progressively reduced accuracy for pixels near the end of the line.

These options can be implemented by appending a control bit to the data streams 13 of Figure 3 in the pipeline so that the operational elements can detect a change of line, with options 2) and 3) taking increasingly complex action on the speed of the calculation hardware, it may also be possible to label pixels with additional control bits as they proceed through the calculation so that occlusion can be detected by labelling corresponding pixels in the data stream from each camera. This will label images of objects at shorter ranges before those at longer ranges, therefore the hardware required for option 3) above could also deal with occlusion (see fig 8) but as the accuracy characteristics would then be a function of the scene as well as of the image position, an additional output indicating the number of cameras on which the pixel correspondence was based would need to be produced.

The accuracy of range estimates from any passive sensor must depend on the scene data. It is only where the scene data is varying that correspondence can be established between pixels in different views. In the proposed implementation of Figure 3 if the variance is plotted against test range it may be shown that an isolated feature would produce a family of variance curves as a function of x as illustrated in fig. 5. This can give problems in the range estimation of isolated features as the minimum obtained is single sided. It also gives problems because the variance is low for all range values when away from the isolated feature and can often be lower than the variance minimum shown at the feature.

A modified algorithm may be devised to subtract the mean variance estimate from the variance profile, it could also reform the data to give a symmetrical minimum by choosing a symmetrical arrangement of cameras with the reference image in the centre. The combined results of these operations shows that for an isolated feature, the minimum variance now has a symmetrical and flat-bottomed profile where the width of the minimum gives a clear indication of the tolerance on the estimated range.

The preceding analysis assumes an idealised sensor array. In a real system the elements of the sensor array will not be perfectly aligned, they will also differ in their overall sensitivity and will not exhibit an exactly uniform response over the whole image field. Additional practical problems are introduced when interfacing the sensor array to the analysis hardware as electrical noise can be picked up on signal lines, and the quantisation levels in the analogue to digital converters are only accurate to say + or - one least

significant bit. The analysis also assumes that the surfaces in the scene exhibit perfect Lambertian radiation properties i.e. the apparant intensity of radiation from a point on a surface does not vary with viewing angle. A combined error model can be drawn to illustrate the interaction of these error terms.

The assumption of Lambertian reflection should be a good approximation for most matt engineering materials such as those in an automated assembly cell provided that the illumination of the cell is sufficiently diffuse, particularly as the range of angles tested by a practical array would be small ($\geq$0.1 radians say). For a more reflective surface, any reflection of objects or lights would be superimposed on the reflectance characteristics of the surface. As this is a linear process the simple amplitude correlation described so far would not give a good variance minimum at either the range of the surface or at this range plus the range from this surface to the reflected object.

The reflected image is added to the normal image of the object but for a matt surface the reflected image will not have sharp edges. The effect of the reflection on the analysis can therefor be reduced by calculating the variance profile of the thresholded first difference images rather than pure intensity images. The threshold level would be chosen empirically to suit the polar reflectance characteristics of the typical objects encountered. This approach has the disadvantage of sharply reducing the number of points in the image for which a range estimate can be made, and may therefore require a further stage which would use the lower level information to grow regions in the depth map which were consistent with the edge

information obtained in the first phase. A further difficulty of using the first difference image is that if a line of sight is tangential to an edge of a curved object then the edge will correspond to different points on the curve from different camera positions. This may cause small errors in the range estimation which cannot be compensated for as might have been possible with the raw data.

For more reflective surfaces the use of the first difference image would allow the variance algorithm to detect the range of significant edges in both the object and in the virtual image formed by reflection. This virtual image could be at any apparent range depending on the curvature of the reflecting surface. Further analysis based on edge effect, range of interest, consistency with the world model or polorisation of reflected light would then be required to eliminate the virtual image range measurements.

The use of the first difference images also compensates for any DC bias between different sensors in the array, or for any slow variation in sensitivity between different regions of the same sensor. Unfortunately it also aggravates any random electrical noise or quantisation effects but as these sources of error are usually of small amplitude, the threshold may be set to remove these terms. In any event they are uncorrelated with the image data, the correlation of multiple images inherent in the analysis should therefore filter out any adverse effects.

Errors due to inaccurate alignment of the sensors can be minimised by the use of solid state cameras which are rigidly connected to each other after alignment calibration. This removes drift in the control electronics as a source of alignment error such

as would have occurred with vidicon tube cameras. Due to the short baseline of the array, the mechanical constraints can be much more robust than is possible with a conventional stereo pair, but mechanical shock and thermal expansion of the mountings cannot be eliminated. For maximum accuracy, it will, therefore, be necessary to automatically update the calibration offsets using feedback from exact object positions as they become known. For less accurate work, as the sensor array is a single unit, calibration by the manufacturers can be sufficient. In either case, manual intervention in setting up for a new application is avoided. With this configuration it should be possible to calibrate the array such that all the sensors are aligned to within plus or minus one pixel error in both the x and y directions. In order to prevent aliasing, the sensors will sample at more than the highest spatial frequency present. The highest possible spatial frequency is given by the point spread function (PSF) of the individual cameras, any misalignment will sample different points on this point spread function. This shows up as an amplitude error which in the worst case is given by the alignment error multiplied by the maximum gradient of the point spread function. This source of error is therefore minimised by defocussing the camera array so that the maximum spatial frequency of interest is the same wavelength as the point spread function.

The analysis hardware can be substantially simplified by reducing the data word length used to represent the image. The use of the first difference image will already have reduced the dynamic range needed. The minimum variation between pixels is defined by the integer resolution of the intensity input, the worst case maximum by

the product of the pixel separation times the maximum gradient of a full amplitude point spread function (PSF). For a wavelength of the PSF equal to ten pixels this corresponds to a reduction of two bits in dynamic range. Empirical observation of typical image statistics could allow larger reductions.

If the expected images are high contrast and the image statistics are well known then the bit resolution can be further reduced down to a binary image, or in the case of a first difference image a two bit (magnitude and sign) image. This format is chosen in preference to the more usual edge detection or zero crossing image as these techniques produce a feature which is one pixel wide, any small amplitude noise in the original image combined with a marginal edge could easily result in an apparent lateral displacement of the edge. This would prevent the correlation algorithm from tracking the edge. A zero-to-one transition defining the edge is less sensitive to this type of small error.

The alternative formulation of the technique with a scanning perspex block avoids many calibration problems. Sensor matching is no longer a problem, and the parallelism of the effective viewpoints is determined by the parallelism of the sides of the perspex block, which can be controlled to close tolerance. The remaining problem is that of determining the angular position of the perspex block at each image position so that the parallax offset can be accurately calculated. Such a technique was used in the following experimental arrangement.

Using apparatus as shown in Figure 4 a sequence of images may be obtained by rotating a parallel sided block of perspex 40 through fixed angular increments in the line of sight of a camera 41. This produces an effective offset (_ x) given by:-

$$x = \frac{T.R.\sin\left(\ \arcsin[\sin\{\ /R\}]\right)}{\cos(\arcsin[\sin\{\ /R\}])}$$

where:- = angular position of the perspex block

T = thickness of the perspex block

R = refractive index of the perspex block (1.498)

An overhead view of a typical test scene is shown in Fig 6. In one test carried out by the inventor this scene consisted of a machine tool cutter 43 and a spring 42 arranged on the surface marked out with a grid for calibration purposes, and in front of a backdrop 44 tiled with acoustic tiles. These objects were set at ranges between 20 and 100 centimetres. The image was digitised using a 512 by 512 by 7-bit frame grabber at 32 equal angular increments of the perspex block 40.

An othogonal image from the experimental data set is shown in Figure 7, printed in a pseudo gray scale on a conventional graphics printer and clearly showing the parallax motion to be measured. The Hough Transform was calculated for a sample line (Fig. 8). The calculation was repeated for each line in the orthogonal image to produce sparse depth maps, as shown in
Figs 10, 11 and 12. An image of the test scene from one camera alone is shown in Figure 9 for comparison with the depth mas shown in Figures 10, 11 and 12. This shows good to sensitivity features which are normal both to the line of sight of the camera and normal to the line of traverse of the linear array.

A second experiment has been constructed where the rotation of the perspex block is co-axial with the line of sight of the camera, producing a circular offset rather than a linear one. This will give equal sensitivity irrespective of feature orientation.

The results obtained so far have shown that a planar array image sensor can avoid some of the ambiguities inherent in the conventional stereo image analysis by the provision of extra information. This allows an analysis algorithm to be used to derive a depth map which is simple, operating on binary data in a single pass. It is, therefore, potentially convenient to implement in hardware.

The limited depth range over which accurate measurements can be achieved with a short baseline sensor can be seen by inspection of the range scale in Figure 8 for the dynamic control task, this should not prove a problem as fine control is usually only required at close range. For the general object recognition task, more time can be allowed, therefore the range information can be used in conjunction with other image analysis techniques - for instance, to provide scale information to constrain the search space of a shape based algorithm.

A limitation of any passive stereo matching procedure is that depth information can only be deduced where there are identifiable features in the image. This effect could be avoided if the scene was illuminated with a projected image similar to structured light, to produce artificial edges in otherwise featureless areas. As the analysis does not use the information of what pattern is projected, or where from, the projection system does not have any expensive requirements for accurate equipment or time consuming set-up procedures.

CLAIMS

1 An automatic machining or assembly system including a comparator for comparing the intensity of a pixel in a first image of a scene produced by a sensor with a corresponding pixel and pixels increasing displaced from the corresponding pixel in a second image of the same scene displaced with respect to said first image and for producing signals representing image depth the magnitude of which is determined by the relative displacement of compared pixels having minimum intensity variation or by a second sensor linearly displaced from the first sensor or by optical defraction means between the first sensor and the scene when rotated to a new position .

2 An automatic machining or assembly system as claimed in Claim 1 and wherein the first and second images are produced by first and second sensors respectively of a plurality of at least three such sensors linearly displaced with respect to each other.

3 An automatic machining or assembly system as claimed in claim 1 and wherein the first and second images are produced by rotatable optical diffraction means between the sensor and the scene when rotated to fist and second position.

4 An industrial vision system including either at least three spaced-apart sensors each sequentially and synchronously producing a separate one of a corresponding plurality of at least three images of a scene, or a single sensor and means for periodically varying its field of view to produce said at least three images frame storage means storing data representing the intensity of a plurality of elements o each image which is produced and comparator means for

sequentially comparing the intensity of an element of an image produced by one sensor with selected elements of an image produced by another sensor and for producing a range signal the magnitude of which is determined by the relative displacement of the elements of each image giving rise to the minimum intensity variation therebetween.

5 A system as claimed in any preceding claim including a regular geometric array of three or more sensors and wherein the comparator or comparator means compares intensities of pixels in the images produced by the sensors relative to a datum pixel in theimage from a datum sensor adjusted by amounts directly proportional to the position of the sensor in question relative to the datum sensor.

6 A system as claimed in any preceding claim and wherein the sensors are television cameras.

7 A system as claimed in any preceding claim and wherein the intensities of pixels in the images produced by the sensors are encoded in binary form and stored sequentially in frame stores for later bit-bit-bit comparison with corresponding encoded pixels produced by other sensors.

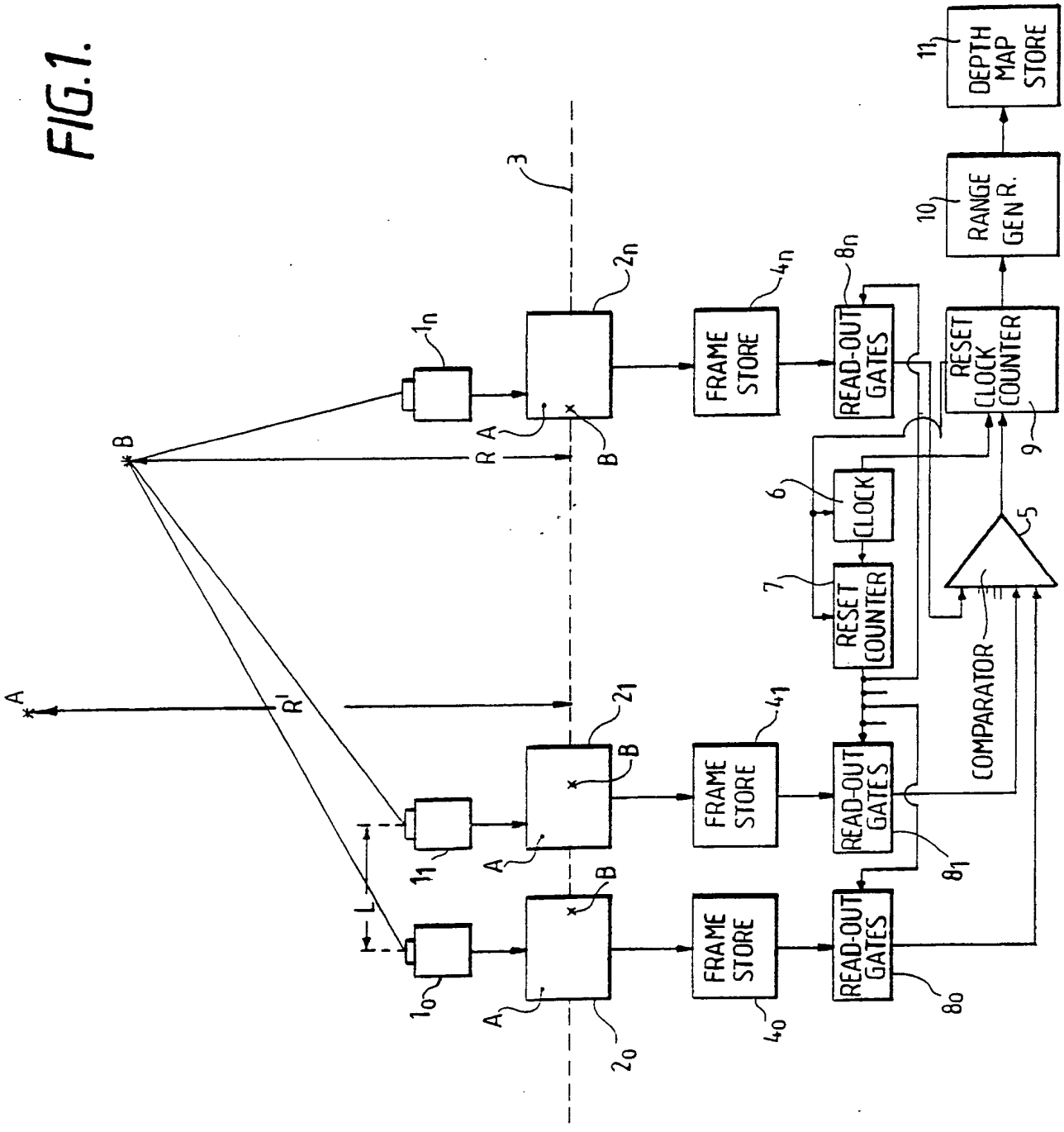
8 A system as claimed in any preceding claim and wherein he intensities of the pixels in each image are transformed using a Hough transform technique into parameter space for processing to find the slope of features in an orthogonal image set.

9 A system as claimed in any preceding claim and wherein signals representing each pixel in each image produced by the sensors are applied to a microprocessor with a program algorithm to produce a corresponding epipolar or orthogonal image.

10 A system as claimed in any preceding claim and wherein a linear array of sensors is used and the comparator or comparison means includes a series of D type registers.

11 A system as claimed in claim 9 and wherein said program algorithm produces range data from pixel intensity variance with pixel offset from each sensors line of sight.

FIG.1.



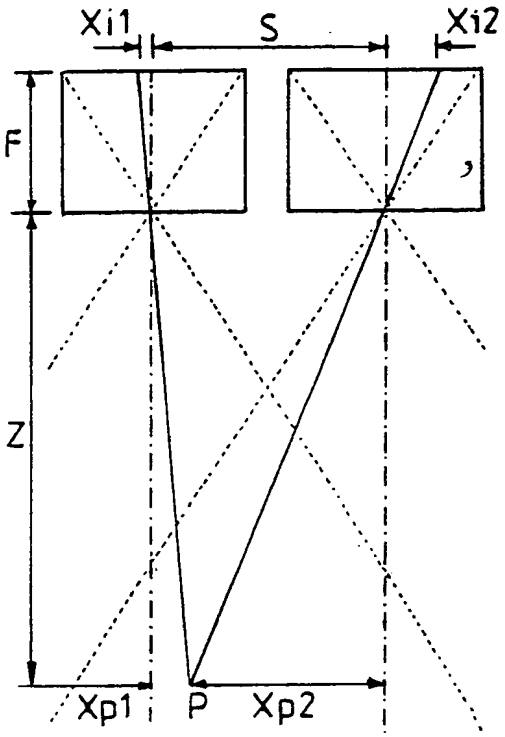


Fig.2.

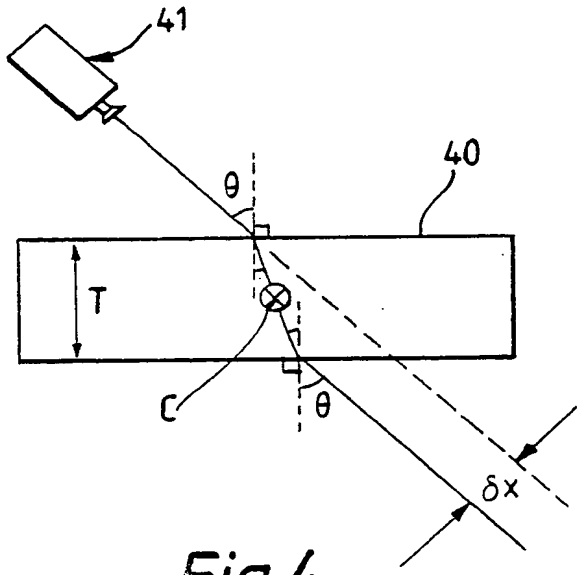


Fig.4

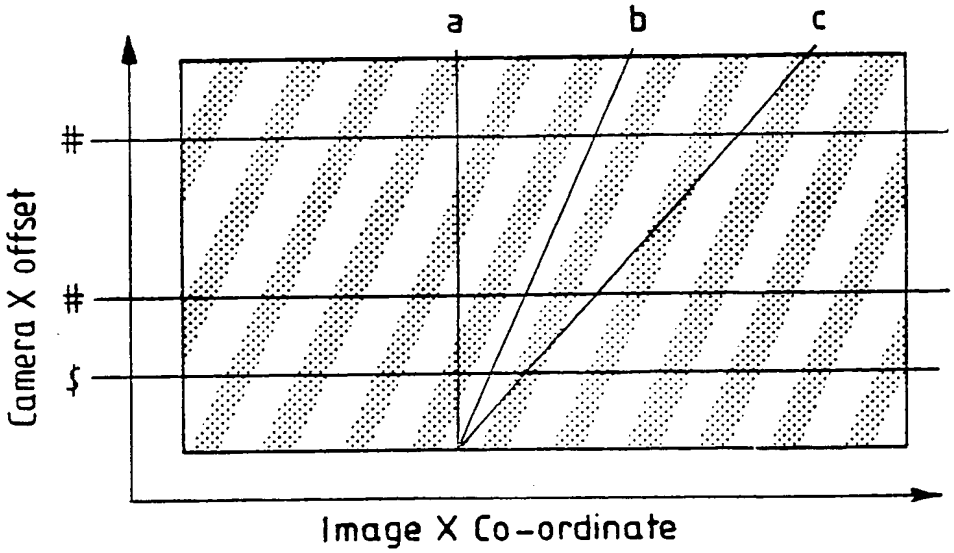


Fig.5.

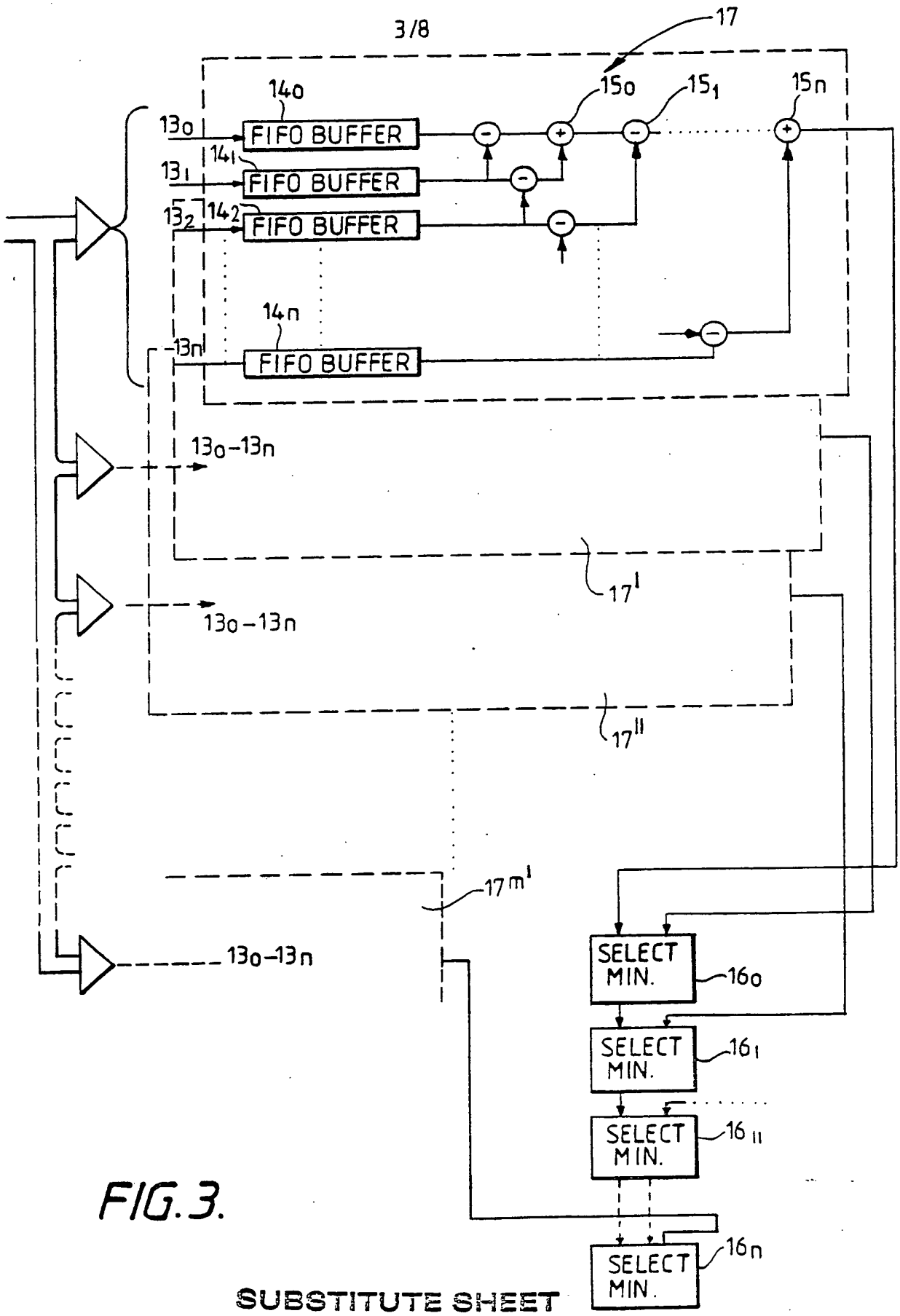


FIG. 3.

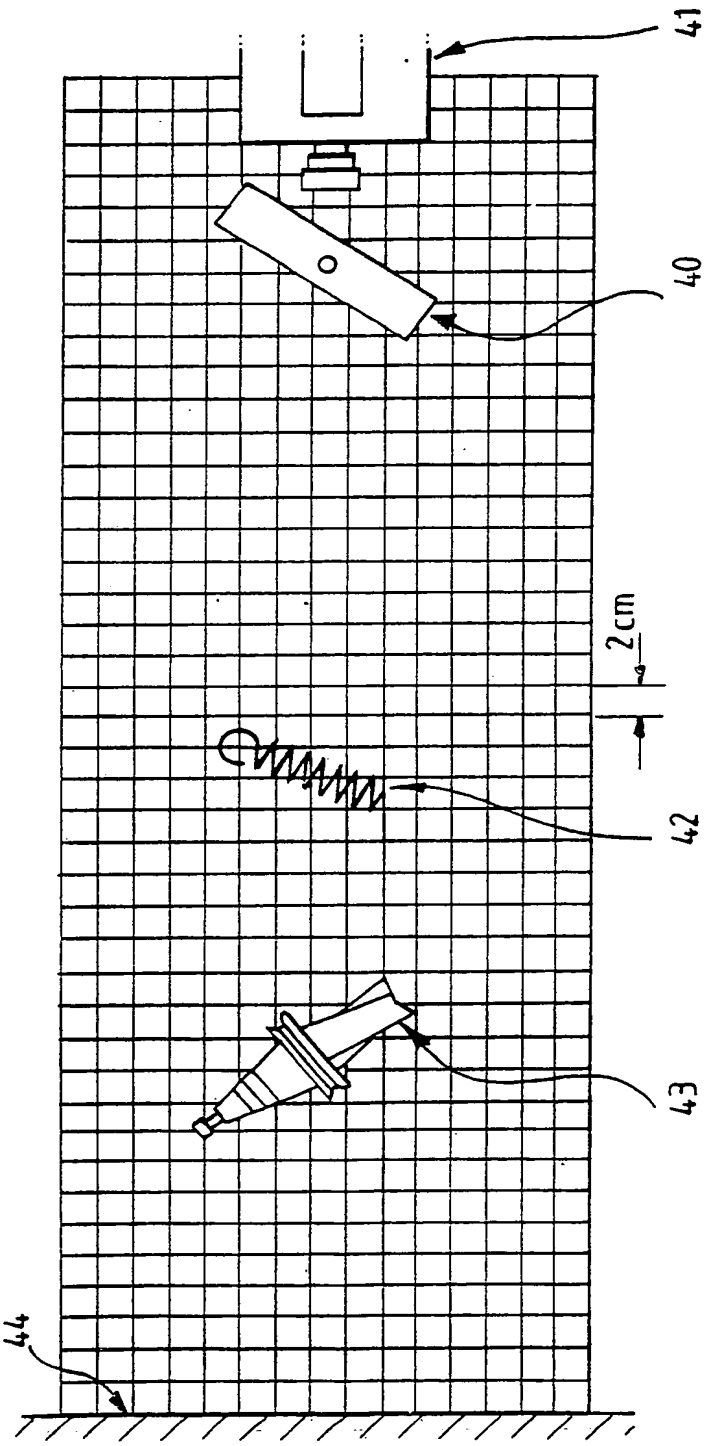


Fig.6.

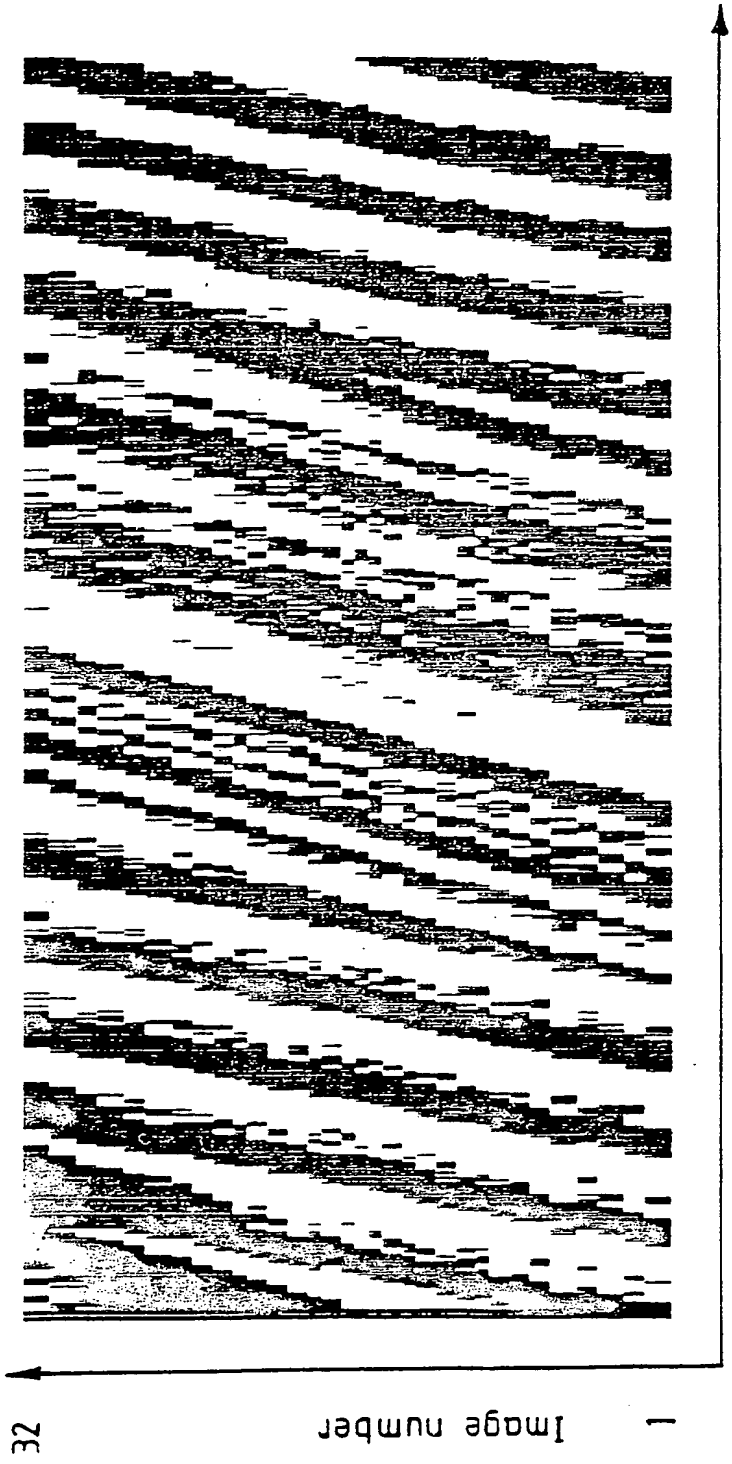


Fig. 7.

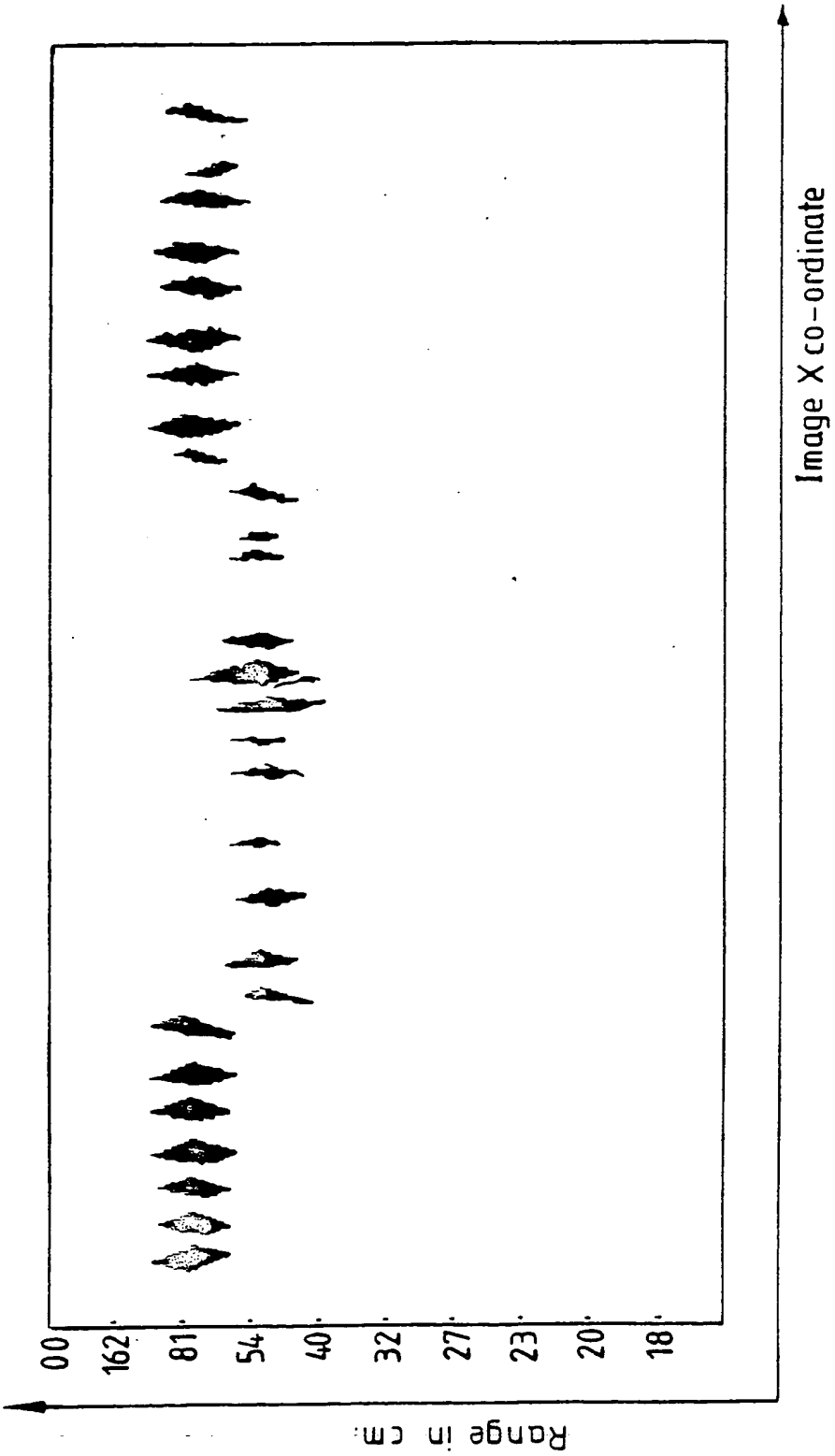


Fig.8.

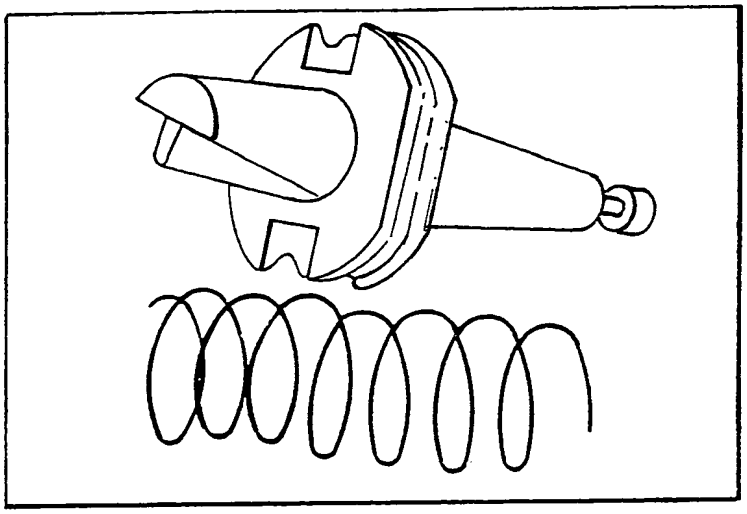


Fig.9.

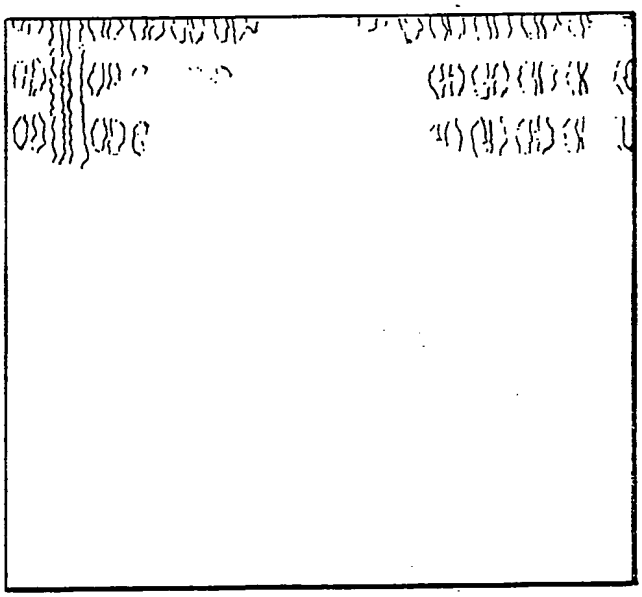


Fig.10.

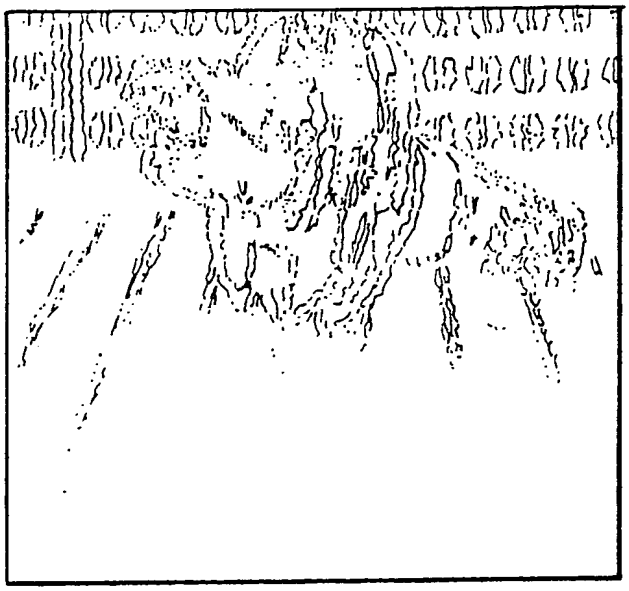


Fig.11.

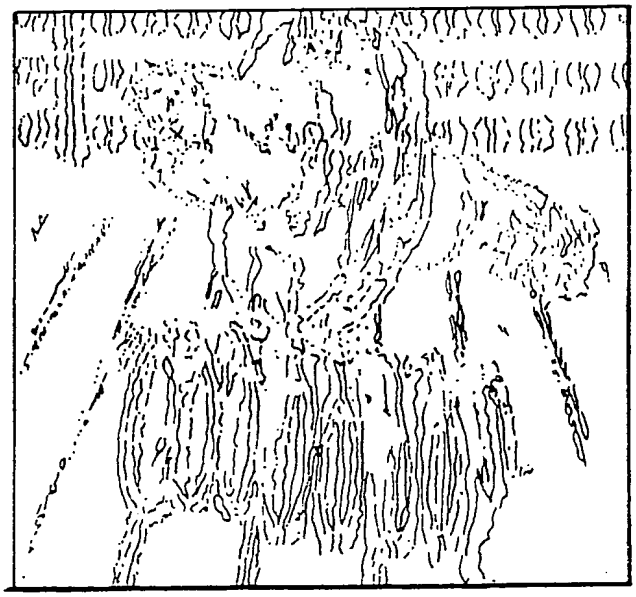


Fig.12.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 :		(11) International Publication Number:
G06F 15/70, G01C 3/00	A3	WO 88/ 02518
		(43) International Publication Date: 7 April 1988 (07.04.88)

(21) International Application Number: PCT/GB87/00700

(22) International Filing Date: 2 October 1987 (02.10.87)

(31) Priority Application Number: 8623718

(32) Priority Date: 2 October 1986 (02.10.86)

(33) Priority Country: GB

(71) Applicant *(for all designated States except US)*: BRITISH AEROSPACE PUBLIC LIMITED COMPANY [GB/GB]; 11 Strand, London WC2N 5JT (GB).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)* : WRIGHT, Steven, M. [GB/GB]; Department of Engineering, Manufacturing Engineering Group, University of Cambridge, Mill Lane, Cambridge CB2 1RX (GB).

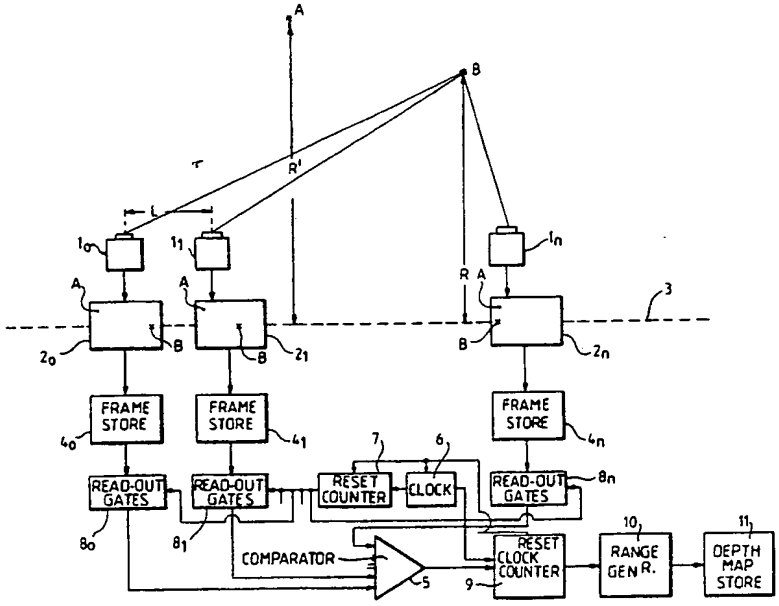
(74) Agent: EASTMOND, John; Corporate Patents Department, British Aerospace PLC, Brooklands Road, Weybridge, Surrey KT13 0SJ (GB).

(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
19 May 1988 (19.05.88)

(54) Title: REAL TIME GENERATION OF STEREO DEPTH MAPS



(57) Abstract

An automatic machining or assembly system including a comparator for comparing the intensity of a pixel in a first image of a scene produced by a sensor with a corresponding pixel and pixels increasingly displaced from the corresponding pixel in a second image of the same scene displaced with respect to said first image and for producing signals representing image depth the magnitude of which is determined by the relative displacement of compared pixels having minimum intensity variation or by a second sensor linearly displaced from the first sensor or by optical diffraction means between the first sensor and the scene, when rotated to a new position.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**FR**	France	**ML**	Mali
AU	Australia	**GA**	Gabon	**MR**	Mauritania
BB	Barbados	**GB**	United Kingdom	**MW**	Malawi
BE	Belgium	**HU**	Hungary	**NL**	Netherlands
BG	Bulgaria	**IT**	Italy	**NO**	Norway
BJ	Benin	**JP**	Japan	**RO**	Romania
BR	Brazil	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SN**	Senegal
CH	Switzerland	**LI**	Liechtenstein	**SU**	Soviet Union
CM	Cameroon	**LK**	Sri Lanka	**TD**	Chad
DE	Germany, Federal Republic of	**LU**	Luxembourg	**TG**	Togo
DK	Denmark	**MC**	Monaco	**US**	United States of America
FI	Finland	**MG**	Madagascar		

INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 87/00700

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC[4]: G 06 F 15/70; G 01 C 3/00

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
IPC[4]	G 06 F 15/62; G 01 C; G 06 K 9/64

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category [10]	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
X	Proceedings of SPIE - Intelligent Robots: Third International Conference on Robot Vision and Sensory Controls RoViSeC3, Cambridge, Massachusetts, 7-10 November 1983, volume 449, part 1, SPIE - The International Society for Optical Engineerging, (Bellingham, Washington, US), G. Hobrough et al.: "Stereopsis for robots by iterative stereo image matching", pages 94-102 see pages 94-102 and in particular chapters 3 and 4 --	1
A	1985 IEEE International Conference on Robotics and Automation, St. Louis, Missouri, 25-28 March 1985, IEEE Computer Society Press, (US), J. Amat et al.: "A vision system with 3D capabilities", pages 2-5 see pages 2-5 and in particular paragraph 1.3 -- ./.	1-3

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
7th March 1988	13 APR 1988
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	M. VAN MOL

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
A	Proceedings of the Third Workshop on Computer Vision: Representation and Control, Bellaire, Michigan, 13-16 October 1985, IEEE Computer Society, (US), R.C. Bolles et al.: "Epipolar-plane image analysis: a technique for analyzing motion sequences", pages 168-178 see pages 168-178 and in particular figures 1-9 and page 168, chapter: "Epipolar-plane images" --	8,9
A	WO, A, 86/05642 (EASTMAN KODAK CO.) 25 September 1986 see figure 5 --------	3

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. GB 8700700 SA 18988

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 25/03/88
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A- 8605642	25-09-86	EP-A- 0213188	11-03-87

EPO FORM P0479

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

THIS PAGE BLANK (USPTO)